Mechanical Systems and Signal Processing 41 (2013) 510–525

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition

George Georgoulas [a], Theodore Loutas [b], Chrysostomos D. Stylios [a,*], Vassilis Kostopoulos [b]

[a] Department of Informatics and Telecommunications Technology, Technological Educational Institute of Epirus, GR-47100, Greece
[b] Department of Mechanical Engineering and Aeronautics, University of Patras, GR-26500, Greece

ARTICLE INFO

Article history:
Received 11 October 2012
Received in revised form
18 February 2013
Accepted 25 February 2013
Available online 14 May 2013

Keywords:
Bearings
Empirical Mode Decomposition
Hilbert Huang transform
Anomaly detection
One-class classification
Ensemble detector

ABSTRACT

Aiming at more efficient fault diagnosis, this research work presents an integrated anomaly detection approach for seeded bearing faults. Vibration signals from normal bearings and bearings with three different fault locations, as well as different fault sizes and loading conditions are examined. The Empirical Mode Decomposition and the Hilbert Huang transform are employed for the extraction of a compact feature set. Then, a hybrid ensemble detector is trained using data coming only from the normal bearings and it is successfully applied for the detection of any deviation from the normal condition. The results prove the potential use of the proposed scheme as a first stage of an alarm signalling system for the detection of bearing faults irrespective of their loading condition.

1. Introduction

The fault diagnosis of rolling element bearings is very important for improving mechanical system reliability and performance in rotating machinery, since bearing failures are among the most frequent causes of breakdowns in rotating machinery. For example in the case of large induction motors, bearing faults can account for up to 44% present of the total number of failures [1]. When a fault occurs in a bearing, periodic or quasi-periodic impulses appear in the time domain of the vibration signal, while the corresponding bearing characteristic frequencies (BCFs) and their harmonics emerge in the frequency domain [2]. However, in the early stage of bearing failures, the BCFs usually carry very little energy and are often suppressed/hidden by severe noise and higher-level vibrations. Consequently an effective signal processing method is of utmost importance for the extraction of damage sensitive features during the condition monitoring of bearings, especially during the initial fault occurrence.

Until now, in the field of bearing fault detection, a variety of approaches and time–frequency signal processing tools have been utilized. Wavelet transform (WT) has been widely used as a de-noising tool as well as for feature extraction. Loutas and Kostopoulos [3] presented a recent review on the applications of WT in rotating machinery diagnostics and prognostics. Abbasion et al. [4] studied the condition of an electric motor with two rolling bearings (one next to the output shaft and the other next to the fan) with one normal state and three faulty states each. De-noising via the Continuous

* Corresponding author. Tel.: +306974638830; fax: +302681050330.
E-mail address: stylios@teiep.gr (C.D. Stylios).

0888-3270/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ymssp.2013.02.020

Wavelet Transform (CWT) (Meyer wavelet) was conducted and Support Vector Machines (SVMs) were used for the fault classification task. Results have shown 100% accuracy in fault detection. Ocak et al. [5] developed a new scheme based on wavelet packet decomposition and Hidden Markov Models (HMMs) for the condition monitoring of bearing faults. In this scheme, vibration signals were decomposed into wavelet packets and the node energies of the 3-level decomposition tree were used as features. Based on the features extracted from normal bearing vibration signals, an HMM was trained to model the normal bearing operating condition. The probabilities of this HMM were then utilized to track the condition of the bearing.

The Empirical Mode Decomposition (EMD) introduced by Huang et al. [6] is another time–frequency tool, which has found growing applications in various fields. It is a self-adaptive method that does not depend on predefined functions as WT, Wigner Ville, Fast Fourier Transform (FFT) and short-time FFT do, and thus it is more suitable and attractive for the analysis of highly non-linear, non-stationary signals as those deriving from the vibrations of rotating machinery. EMD decomposes a signal into a number of basic constituent signals called Intrinsic Mode Functions (IMFs). Lei et al. [7] suggested a method relying on Wavelet Packets Transform (WPT) and Empirical Mode Decomposition (EMD) to pre-process vibration signals and extract fault characteristic information. Each of the raw vibration signals was decomposed with WPT using Daubechies wavelets with 10 vanishing moments "db10" at level 3. From the plethora of features extracted at each sub-band, the most relevant ones were selected via distance evaluation techniques and forwarded into a Radial Basis Function (RBF) network to automatically identify different faults (inner race, outer race, roller) in rolling element bearings.

The same research group [8], implemented a very thorough approach through an efficient damage identification scheme, which included the EMD decomposition of a vibration signal. The signal was fully decomposed into IMFs via the EMD algorithm. Then a complicated data fusion procedure via feature extraction, decrease of input vector dimensionality and, finally, pattern recognition/classification was implemented. The results effectively classified different damage conditions. However, the used number of IMFs was too large and, as a result, the whole algorithm was overly complex.

Yu et al. [9] also applied EMD on vibration signals collected from piezoelectric transducers. The experiments involved roller bearings with local faults. The bearing vibration signal was collected by a piezoelectric transducer and sampled at 4096 Hz. This frequency was adequate in order to capture the modulated pulse force of the outer bearing race fault and its first harmonic (76 and 150 Hz, respectively). The vibration signal was de-noised via WT implementation (db10 wavelet) and was fully decomposed into IMFs with the EMD algorithm. The Hilbert spectrum gave the instantaneous frequency of each IMF. Based on the Hilbert spectrum the appropriate IMFs were picked in order to evaluate the roller bearing fault evolution (outer race). The local marginal spectrum from the selected IMFs effectively identified the spectral lines of the fault. The same efficiency was observed for inner race faults also. In this case, the EMD method was proven superior to the classical envelop spectrum method for the bearing fault identification.

Peng [10] investigated the effectiveness of EMD for analyzing the non-stationary cutting force signal of machining processes. Towards the effective detection of tool breakage, the Hilbert spectrum was utilized as well as the energies of the characteristic IMFs associated with characteristic frequencies of the milling process.

Guo and Tse [11] studied the ensemble EMD (EEMD), a variation of classical EMD, which eases the problem of mode mixing, - in real vibration signals generated from defective bearings. They performed a series of investigations to reveal the relationship between the amplitude of the added white noise and the number of ensemble members for the minimization of the mode mixing problem and concluded that a higher number of ensemble members leads to smaller RMS error.

Žvokelj et al. [12,13] proposed the EEMD-based Multiscale Principal Component Analysis (EEMD-MSPCA) and the EEMD-based Multiscale Kernel PCA (EEMD-MSKPCA) techniques to overcome the non-adaptive nature of WT utilized in conventional MSPCA. They applied their techniques in vibration as well as acoustic emission recordings from large-size slow-speed bearings and highlighted the improvement of signal-to-noise-ratio and the enhanced diagnostic capability. The only potential drawback is that the proposed techniques are computationally intensive.

Most of the work involving bearing fault detection relies on information coming both from the normal as well as the faulty class. As a matter of fact among the approaches reported in the aforementioned literature only the methods proposed by Ocak et al. [5] and by Žvokelj et al. [12,13] were based solely on information coming from the normal class. However in most real life applications, data from the possible faulty modes are not readily available, making the binary (or multi-class) classification approach very difficult and impractical.

The main innovation of this article stems from the proposal of a combined and integrated anomaly detection approach to bearing fault detection. The anomaly detection is based on a hybrid or multi-strategy ensemble detector; it is the first time that an ensemble approach is applied in the bearing fault detection field. The proposed scheme is successfully applied to detect faults based on a small number of features extracted using EMD. Nowadays, EMD is a state-of-the-art signal processing method [6] that has been successfully employed in a number of fault detection studies in electrical [14] and mechanical systems [7–13]. This work shows that the suggested integrated methodology for fault detection can be envisioned as the "first line of defence" against incipient faults that can eventually build up to failures of a component or a system, with notable performance while requiring a minimum number of configuration parameters. Moreover the small number of features involved as well as the simplicity of the detectors comprising the ensemble, make this approach suitable for online implementation.

The rest of the paper is structured as follows: Section 2 summarizes the theory of EMD. The anomaly detection approach, as well as the three individual anomaly detectors employed, is presented in Section 3. In Section 4 the proposed

novel fault detection approach is described in detail. Multiple experimental results that prove the efficiency of the proposed scheme are presented in Section 5. Finally, the conclusions and objectives for further research investigation are drawn in Section 6.

2. Empirical Mode Decomposition and the Hilbert Huang spectrum

Most real life processes are inherently non-linear and non-stationary. Therefore, using techniques that assume linearity and/or stationarity no matter how solid their mathematical background is, can lead to suboptimal and misleading results or even to results that have completely no connection to the physical system that they are supposed to "model". EMD and the Hilbert Huang transform, introduced by Huang et al. [6] in 1998 bridge this gap between theory and real life.

EMD lacks rigorous mathematical analysis and it decomposes the signal into a collection of IMFs, where an IMF represents a simple oscillatory function with the following conditions that have to be satisfied:

1. The number of zero crossings and the number of local extrema are equal or they differ by one.
2. The local average (defined by the average of local maximum and local minimum envelops) is equal to zero.

These two conditions practically imply that for an IMF all its local maxima are positive and all its local minima are negative.

The well behaved Hilbert transforms of the IMFs give an alternative approach to time-frequency decomposition which results from the traditional short time Fourier transform and the most recently developed wavelet transform.

More specifically, given a signal $x(t)$ the EMD algorithm can be summarized as follows:

1. identify all local minima and local maxima of the given signal ($x(t)$).
2. create an upper ($e_{\max}(t)$) and a lower ($e_{\min}(t)$) envelope interpolating between successive local maxima and local minima, respectively (usually via cubic interpolation).
3. calculate the running mean $m(t)=e_{\min}(t)+e_{\max}(t)/2$
4. subtract the mean from the signal to extract the detail $d(t)=x(t)-m(t)$.
5. repeat the whole process replacing $x(t)$ with $m(t)$ until the final residual becomes a constant value, a monotonic function or a function with only one extremum from which no more IMFs can be extracted (or a user specific number of IMFs has been extracted—application dependent).

In practice, step 4 may not produce a valid IMF. As a result sifting needs to take place, which implies the iteration of steps 1 to 4 upon the detail $d(t)$ until a specific certain criterion is met. A number of criteria have been proposed in the literature for stopping the sifting process [15] (i.e. Cauchy type criterion [6,16], S-number criterion [17], fixed sifting time criterion [18], mean value criterion [19], orthogonality criterion [20], bandwidth criterion [21], the energy difference tracking method [22], etc).

In this work we employed a variant of the criterion proposed by Rilling et al. [23] that is:

Iterate the sifting process until $\sigma(t)<\theta_1$ for a fraction $(1-\alpha)$ of the total duration and $\sigma(t)<\theta_2$ for the remaining fraction, where $\sigma(t)=|m(t)/a(t)|$ and $a(t)=e_{\max}(t)-e_{\min}(t)/2$ (referred as the mode amplitude [23]). The employed variant – implemented in [24] – uses only θ_1 and α leading usually to a more robust implementation without compromising the efficiency of the algorithm. A recommended set of values, which were in fact used in all the experiments, is $\theta_1=0.05$ and $\alpha=0.05$.

After the completion of the aforementioned procedure, the original signal $x(t)$ is eventually decomposed into a sum of IMFs plus a residual term:

$$x(t)=\sum_k IMF_k(t)+r(t) \tag{1}$$

Following the EMD, the Hilbert transform can be applied to each IMF separately. The Hilbert transform $y(t)$ of any signal $x(t)$ is defined as [25]:

$$y(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(t)}{t-\tau}d\tau \tag{2}$$

Utilizing $y(t)$ and $x(t)$ the associated analytical signal $z(t)$ is defined as:

$$z(t)=x(t)+iy(t)=a(t)e^{i\theta(t)} \tag{3}$$

where $a(t)=\sqrt{x^2(t)+y^2(t)}$ is the envelope of the signal and $\theta(t)=\arctan(y(t)/x(t))$is the instantaneous phase. The instantaneous frequency can be calculated as the derivative of the instantaneous phase:

$$\omega(t)=\frac{d}{dt}\theta(t) \tag{4}$$

By performing the Hilbert transform on each IMF the original signal can be expressed as the real part (RP) in the following form:

$$x(t) = RP\left(\sum_j a_j(t)e^{i\theta_j(t)}\right) = RP\left(\sum_j a_j(t)e^{i\int \omega_j(t)dt}\right) \quad (5)$$

The above equation gives both the amplitude and the frequency of each component as a function of time. This time–frequency distribution of the amplitude is called the Hilbert–Huang (HH) spectrum ($H(\omega,t)$). Fig. 1 depicts the decomposition process of a vibration signal from a bearing without fault and Fig. 2 its corresponding HH spectrum. Figs. 3 and 4 present the IMFs and the HH spectrum for a bearing with a ball fault; all figures were created using the

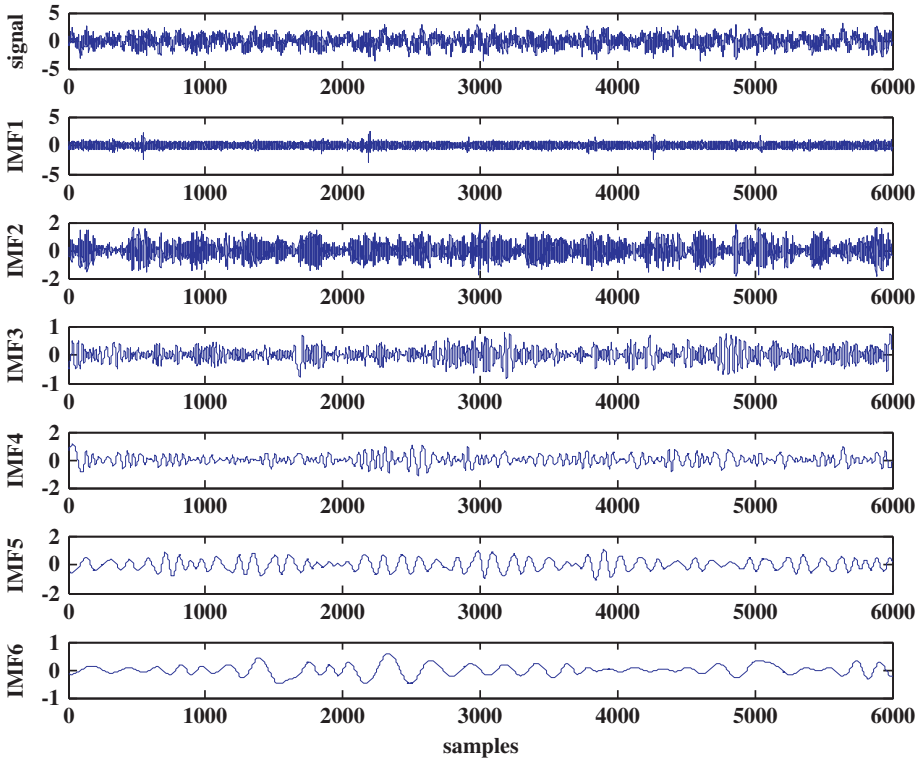

Fig. 1. Original vibration signal and the first six produced IMFs of a healthy bearing.

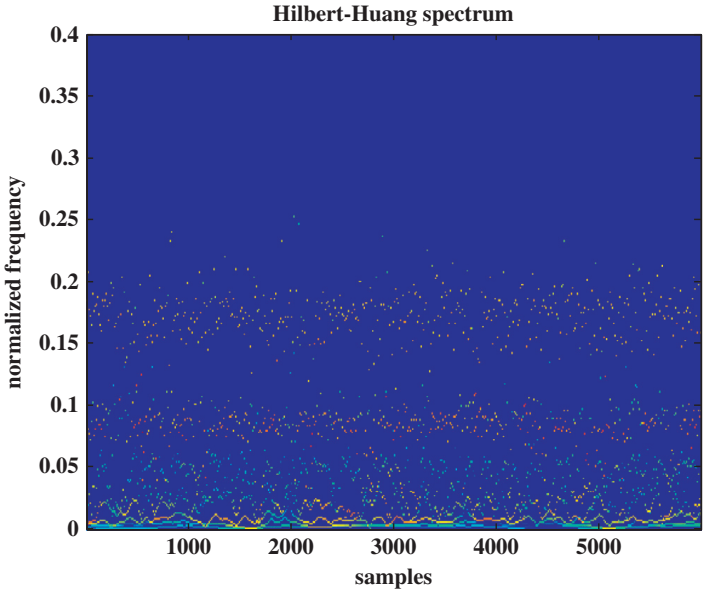

Fig. 2. The HH spectrum of the original vibration signal depicted in Fig. 1 (created using the EMD toolbox [24]).

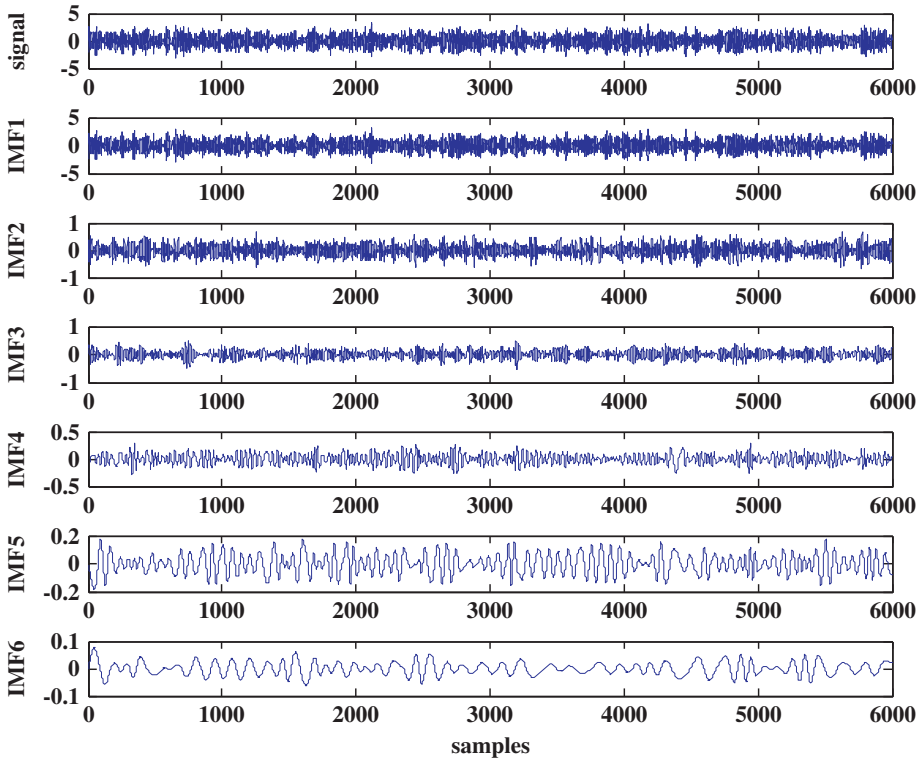

Fig. 3. Original vibration signal and the first six produced IMFs of a bearing with ball fault.

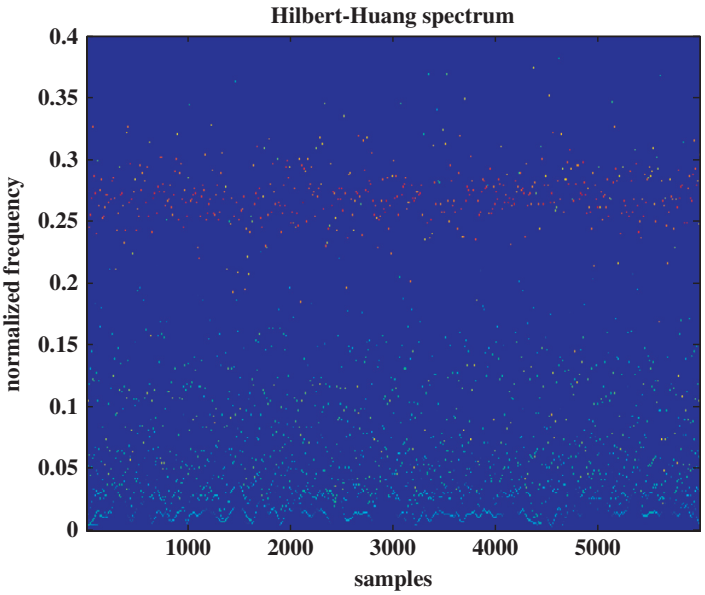

Fig. 4. The HH spectrum of the signal depicted in Fig. 3. (created using the EMD toolbox [24]).

EMD toolbox [24]. By a quick inspection of the two HH spectrums it is obvious that the fault causes higher instantaneous frequencies to emerge whereas the inspection of the IMFs reveals that the spread of IMFs' values is also affected. A more structured and justified approach for the selection of the appropriate IMFs is presented in Section 4.

It should be noted that in all subsequent figures, 0.5 of the normalized frequency corresponds to half the sampling frequency (sampling frequency is equal to 12 kHz, see Section 4.1).

3. Anomaly detection

In real life problems, it is sometimes difficult to have access to data regarding all the possible classes/damage modes that a complex engineering system is likely to encounter. This happens because the available data of some of the classes are quite rare or because it is too expensive to collect them. Moreover new classes/damage modes might emerge, i.e.

new unforeseen faults during the operation of a component or a system. Therefore in such cases it is more practical to design a detector instead of a classifier. These kinds of detectors are usually trained using data coming from only one known class (usually the healthy one) and actually detect any object that deviates from what has been learnt before. In the machine learning literature, these detectors are called novelty detectors [26,27] Since information from only one class is exploited, this kind of novelty detectors is closely related with the one-class classification approach [28]. If the new data deviate from a well-defined notion of normal behavior the process is referred to as anomaly detection [29].

Anomaly detection algorithms characterize the normal behavior of a system by sampling baseline data [30]. They are considered to be the first "line of defense" of any condition monitoring system, as in the beginning it should be identified if there is something wrong, then try to identify which fault has occurred and finally, if possible, estimate the remaining useful life of the system. This work is going to focus only on the anomaly detection process of a bearing element in an attempt to provide continuous monitoring and a reliable alarm mechanism for similar systems.

It has been acknowledged in practice that for the quite challenging tasks of novelty and/or anomaly detection, there is no single best model and the success depends on the properties of the data handled. Over the years a number of novelty detection methods have been developed and the interested reader could refer to any of the excellent surveys that can be found in the literature [26,27,29].

Moreover, in pattern recognition problems it is common to combine various experts with the aim of compensating the weakness of each single expert [26]. This combination also known as ensemble learning, is one of the standard and most important techniques for improving classification accuracy in machine learning [31–35]. The underlying concept is to combine a set of models in order to obtain a better composite global model, with more accurate and reliable decisions compared to those provided by using a single model [35]. What is interesting is that the idea of combining individuals "opinions" in order to reach a final decision is humans' "second nature" before making any crucial decision [36].

Here, this procedure is exploited by utilizing a hybrid approach that involves an ensemble of anomaly detectors trained on different representations of the hypothesis space, combined using a simple majority voting scheme. The anomaly detectors in this study are based on the one-class classification approach; in other words the anomaly detection algorithm learns a discriminative boundary around the normal instances using a one-class classification algorithm [29].

To be more specific, the following detectors were employed: (a) a Gaussian anomaly detector which belongs to the family of density methods, (b) a nearest neighbor (NN) anomaly detector which belongs to the family of boundary methods and finally (c) a Principal Component Analysis (PCA) anomaly detector, which belongs to the family of reconstruction methods [28]. Each of the detectors acts upon a number of feature sets as it will be described in the following subsections. The results suggest that this combination of multiple input sets along with multiple novelty detectors achieves the best performance.

One of the most significant reasons behind the selection of these specific detectors is the minimum number of parameters that the end user needs to provide. Among the three, only the PCA anomaly detector requires one parameter which was selected based on an empirical criterion as explained in Section 3.3. It is worth noting that minimizing the number of parameters is one of the major requirements of any detection scheme [26].

3.1. Gaussian anomaly detector

A Gaussian anomaly detector is the simplest member of the density based anomaly detection family. A density based method builds a boundary around the training data (the normal operation data) setting a threshold on the estimated density [28,37]. The Gaussian anomaly detector, as it is declared by its name, assumes that the training data are normally distributed and uses the Gaussian density function to estimate the probability of a given data point $\mathbf{x}$:

$$p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|1/2}\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{6}$$

where d is the dimension of the input space, $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\Sigma}$ is the covariance matrix which are estimated using the training data.

The method is very simple and it imposes a strict unimodal and convex density model on the data. Even if the data are not normally distributed, as in our case, this anomaly detector method is one that is worth trying due to its simplicity [38].

3.2. Nearest neighbor anomaly detector

The NN method belongs to the family of boundary methods. Boundary methods try to "draw" a boundary around the known class without resorting to the intermediate step of density estimation which – as it has been pointed out – is more difficult than the original problem of assigning labels to objects [39,40].

The NN anomaly detector is based on the assumption that normal data instances occur in dense neighborhoods, while anomalies occur far from their closest neighbor [29]. To quantify that, a distance (or similarity) measure D is needed. In the

specific implementation of the NN anomaly detector [24], a new object **x** is classified based on the (Euclidian) distance to its nearest neighbor **y** normalized by the distance between that object (**y**) and its nearest neighbor **z**:

$$D = \frac{||\mathbf{x}-\mathbf{y}||}{||\mathbf{y}-\mathbf{z}||} \tag{7}$$

3.3. Principal Component Analysis anomaly detector

The PCA anomaly detector belongs to the family of reconstruction methods which assume that a number of prototypes or subspaces are capable of describing the normal set and the process of defining these prototypes or subspaces is through a minimization process of the reconstruction error (the difference between the original data and their reconstructed version using the reduced representation). The anomalous situations should be represented worse compared to the normal data and their reconstruction error should be larger.

PCA assumes that data can be embedded into a lower dimensional linear manifold. PCA linearly transforms the original space [41–44] by projecting the d—dimensional data onto the l ($l \leq d$) eigenvectors of their covariance matrix corresponding to the l larger eigenvalues. The procedure is as follows:

a) compute the mean value for each feature and then subtract this mean.
b) calculate the covariance matrix of the zero-mean data matrix, its eigenvalues and its corresponding eigenvectors.
c) retain the eigenvectors corresponding to the l largest eigenvalues and project the input vectors on them to get a reduced representation.

By defining the $d \times l$ matrix **W** whose columns are the retained eigenvectors, then the reconstruction error e_r of a new object is the squared distance between the original object and its mapped version:

$$e_r = ||\mathbf{x}-(\mathbf{W}\mathbf{W}^T)\mathbf{x}||^2 \tag{8}$$

In the case of the PCA anomaly detector, the number l of retained eigenvectors should be provided (the other two anomaly detectors described above do not require any parameter setting). There are a number of approaches for selecting the number of retained eigenvectors [43,44]. Among the various approaches the simplest one is based on the cumulative percentage variance (or the proportion of variance accounted for). Even though there is not a golden rule regarding the threshold imposed on this quantity, a value of 90% can be considered a valid value [12,13,43]. However, this value is always application dependent so supervision by the user is advised.

In this specific application the fraction of the proportion of variance accounted for by the first principal component is $\sim$92%. As a result only one principal component ($l=1$) was retained.

Note: In all three detectors, the threshold that was set to discriminate between normal data and anomalies was set empirically based on the training data and by requiring a rejection rate of the normal data equal to 1% [38]. In the case of the Gaussian detector the threshold could have been set theoretically, however due to the finite number of training examples the empirical approach was also selected [28].

4. Hybrid anomaly detection procedure

The anomaly detection scheme is composed of a number of stages and the whole procedure is shown schematically in Fig. 5. Each stage is described in the subsequent sections along with the employed data.

4.1. Data set

The data used in this research work come from two bearings installed in a motor driven mechanical system [45], one at the drive end of the motor and the other at the fan end. In both bearings three types of faults (outer race, inner race and ball faults) were introduced using electro-discharge machining with various fault diameters. For the case of the outer race faults, experiments were conducted for both fan and drive end bearings with outer raceway faults located at 3 o'clock (directly in the load zone), at 6 o'clock (orthogonal to the load zone), and at 12 o'clock. Each bearing was tested under four different loads, 0, 1, 2, and 3 hp.

For each test, vibration data were collected using accelerometers placed at the 12 o'clock position at both the drive end and fan end of the motor housing. In this study, we employed all available data (not all combinations of fault sizes and loadings are available) coming for all three fault types for both locations (drive end, fan end), for all loadings and for fault sizes 0.007–0.028 in., collected by the accelerometer located at the drive end at 12,000 samples/s. The data corresponding to the drive end accelerometer were selected because they cover a broader set of configurations compared to the fan end accelerometer data.

A more detailed description of the experimental set-up and the apparatus involved can be found in the Case Western Reserve University's website [45].

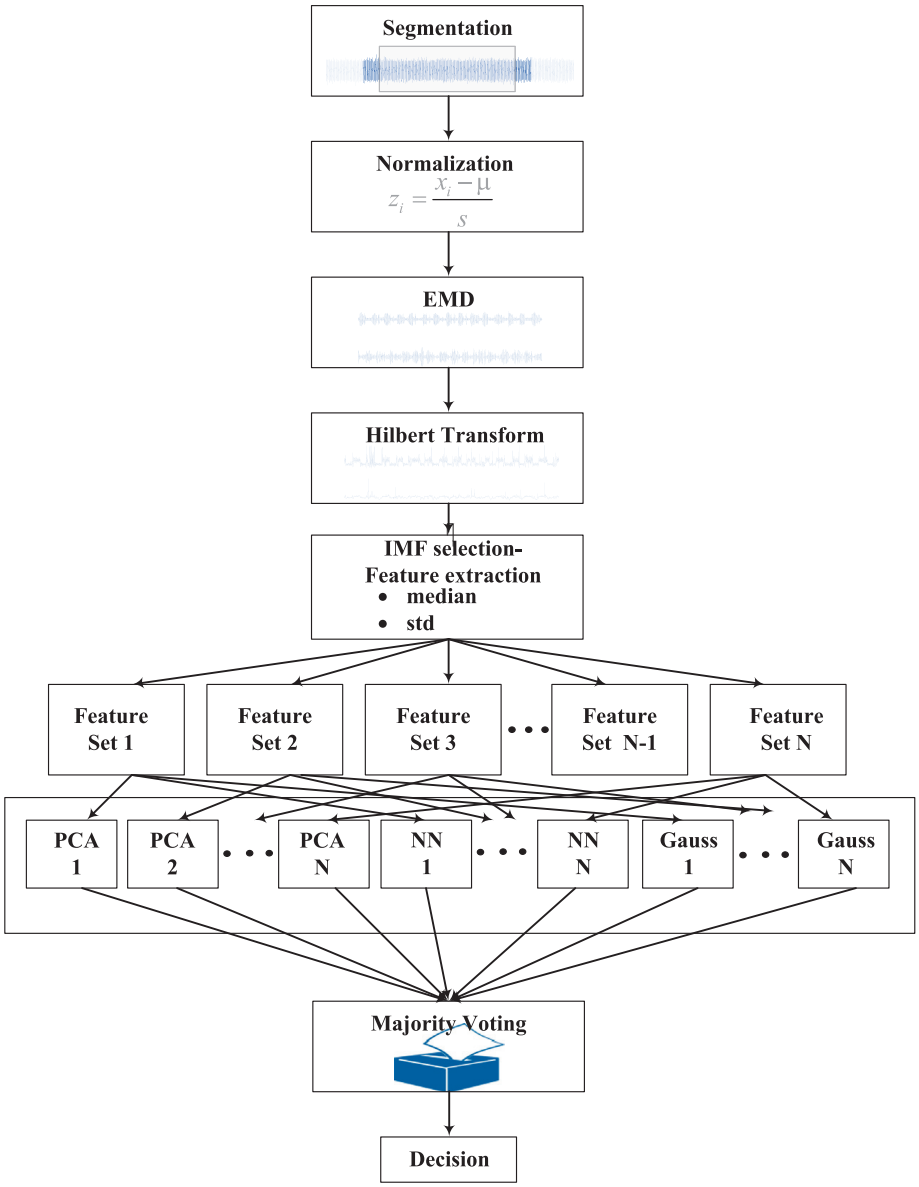

Fig. 5. The hybrid/multistage detection process.

4.2. Segmentation

The detection process is based on the analysis of vibration signals coming from the accelerometers mounted near the drive end. Since the rotation speed is approximately constant the features employed are barely affected by the signal's duration. Therefore instead of segmenting the vibration signal based on the number of complete revolutions of the motor, we used a fixed window of 0.5 s which corresponds to 6000 samples (and approximately ~14.5 revolutions). Therefore from each record we extracted more than one segment without overlapping. The total number of segments extracted is summarized in the following Table 1. Note: segments coming from the same malfunctioning part were put together irrespectively of the fault dimension and the loading condition.

4.3. Normalization

Following the pre-processing step proposed in [46] we normalized the raw vibration signals in order to have comparable magnitudes:

$$z_i = \frac{x_i - \mu}{s} \quad \text{for} \quad i = 1,2,\ldots,N \tag{9}$$

where N is the number of samples of the signal, μ is the estimated mean and s is the estimated standard deviation of the raw signal x, and z is the resulting normalized signal.

Table 1
Summary of extracted segments.

	No. of segment
Normal	280
Inner race fault (DE)	320
Outer race fault (DE)	560
Ball fault (DE)	240
Inner race fault (FE)	239
Outer race fault (FE)	340
Ball fault (FE)	240

Note: DE stands for drive end and FE stands for fan end.

4.4. Feature extraction using EMD

As in conventional classification, in the case of anomaly detection the detector usually does not act upon the signal but on a feature space of much lower dimension than the original input space. This feature extraction process usually improves generalization [47]. On the other hand sometimes feature extraction can be more of an art than a science unless some expert knowledge does exist in order to guide the process [47]. In the case of rolling elements the fault causes changes both in the frequency as well the spread of the IMFs as it can be seen in Figs. 1–4.

Therefore in this study the instantaneous frequency of the first few IMFs coming from the EMD of the vibration signal as well as their spread were selected. Our choice was based on the observation of the IMFs and the HH spectrum during a preliminary set of experiments involving a ball fault. As it can be seen in Fig. 4 the ball fault seems to invoke higher frequency components compared to normal bearing vibration data (Fig. 2). Moreover the spread of the first IMF is increased whereas the spread of the next 2 IMFs is decreased (Figs. 1 and 3). In our preliminary experimentation, we arbitrarily selected only the first two IMFs and the results were satisfactory. However this approach is subjective and might not work the same way under different sampling frequency or/and under different noise conditions. Therefore a more systematic way is required to determine the most representative IMFs for extracting the bearing fault condition information out of vibration signals. A criterion for selecting the two most appropriate IMFs was proposed by Žvokelj et al. [12,13] based on kurtosis and this is in accordance with our intention to come up with as few as possible feature components and as we explained above our preliminary results suggested that two IMFs, although not optimally selected, are sufficient for the specific problem at hand. As it was pointed out the most appropriate IMFs are those of highest impulsive nature, and a way to quantify is by the use of kurtosis:

$$Ku = (1/M)\sum_{i=1}^{M}\frac{(x(i)-\bar{x})^4}{RMS^4} \tag{10}$$

where *RMS* denotes the root mean square value and $\bar{x}$ is the mean value of the discrete time signal $x(i)$ with an M number of samples.

Based on that criterion, we selected the first and the third IMFs (Figs. 6 and 7) and we focused on their instantaneous frequencies (Figs. 8 and 9) resulting by the application of the Hilbert transform as well as their spread.

Apparently the "level" of the instantaneous frequency is different in the case of the normal and in the case of a bearing with a ball fault. A way to quantify this difference is by using its mean value. However, since the estimation of the instantaneous frequency is quite spiky (Figs. 8 and 9), the median [48] was preferred over the more conventional mean value (or a combination of Gaussian smoothing with the estimation of the mean). Moreover the standard deviation of the two IMFs was also included in the feature vector (to quantify their spread).

Consequently four features were obtained:

- f1-median of the instantaneous frequency of the first IMF.
- f2-median of the instantaneous frequency of the third IMF.
- f3-standard deviation of the first IMF.
- f4-standard deviation of the third IMF.

The discriminating capability of the selected features for the problem at hand can be stressed by observing the resulting histograms for the normal and the anomaly/faulty case depicted in Fig. 10. As it can be seen the median of the instantaneous frequency looks more appropriate for the discrimination between the two classes. However the other features also contain complementary information that leverages the detection procedure. To summarize, even though the specific set of features has not been selected using an optimal selection process (through the exploitation of a richer feature bank), it led to perfect detection accuracy within the framework of the ensemble anomaly detector not only for the case of ball faults but for all six type of faults encountered in this experimental set-up.

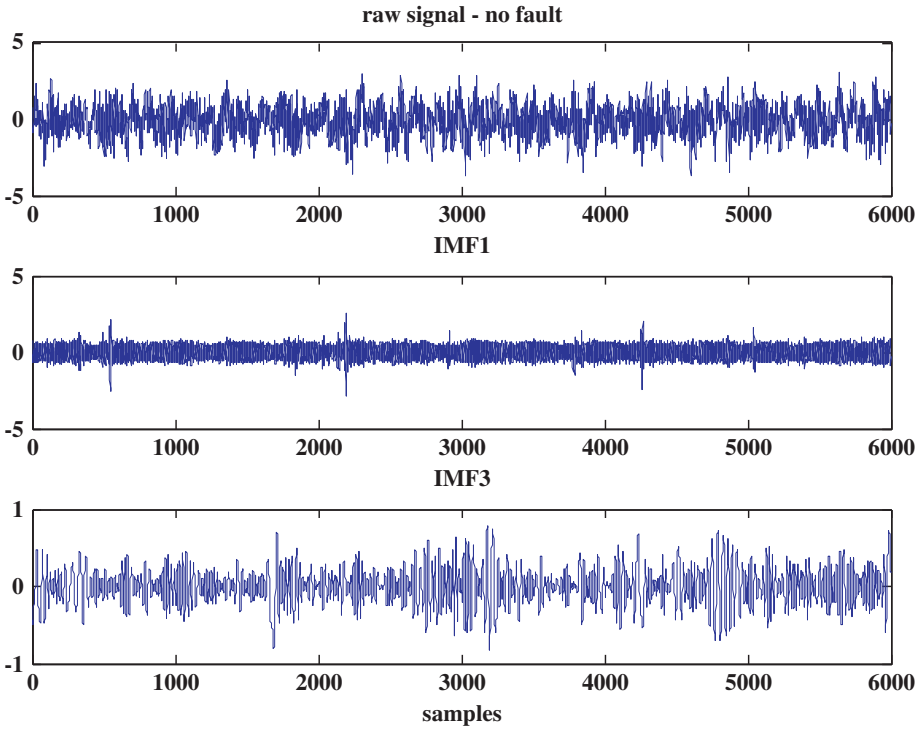

Fig. 6. Raw vibration signal and IMFs 1 and 3 for normal bearing, 2 hp loading.

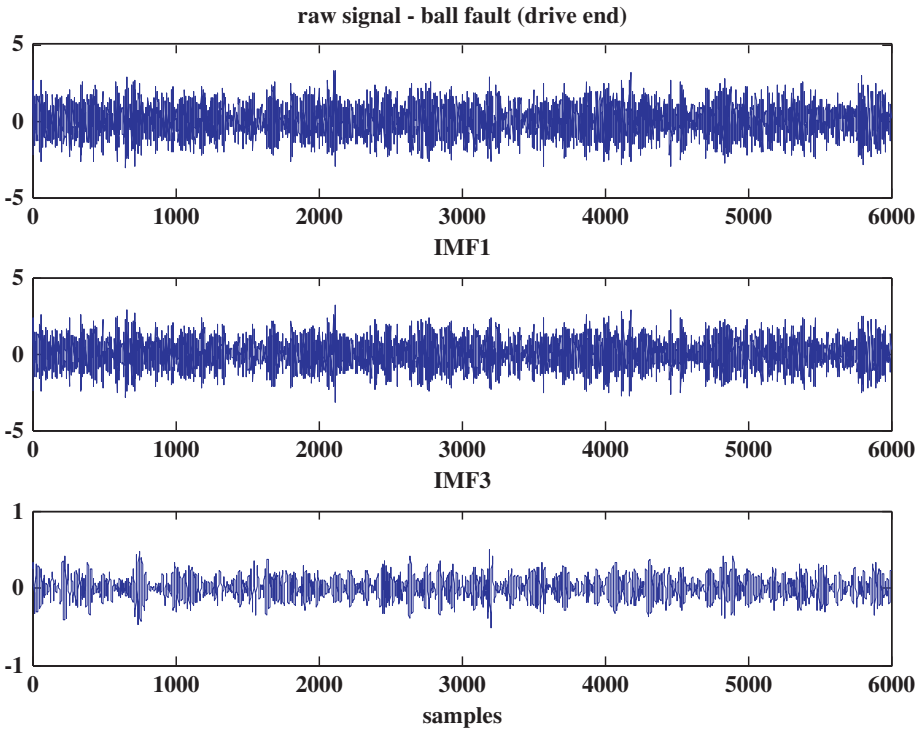

Fig. 7. Raw vibration signal and IMFs 1 and 3 for a bearing with a 0.007 inch ball fault, 2 hp loading located at the drive end.

4.5. Ensemble of anomaly detectors

In this work the anomaly detection is performed by combining the decisions of three different anomaly detectors trained using different sets of input features and combining their decisions through a majority voting scheme which some time is referred to as the plurality vote or the basic ensemble method [32]. Majority voting is probably the simplest way of combining classifiers. However, when accuracy is approximately balanced across the involved methods, majority voting performs better than any more complex combination rule [49].

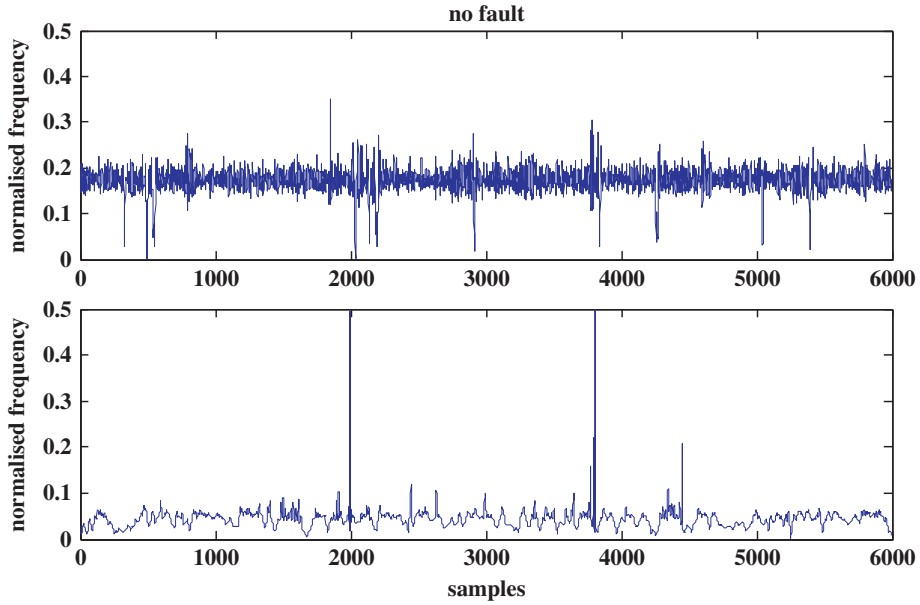

Fig. 8. Instantaneous frequency for the 2 IMFs depicted in Fig. 6, for a normal bearing.

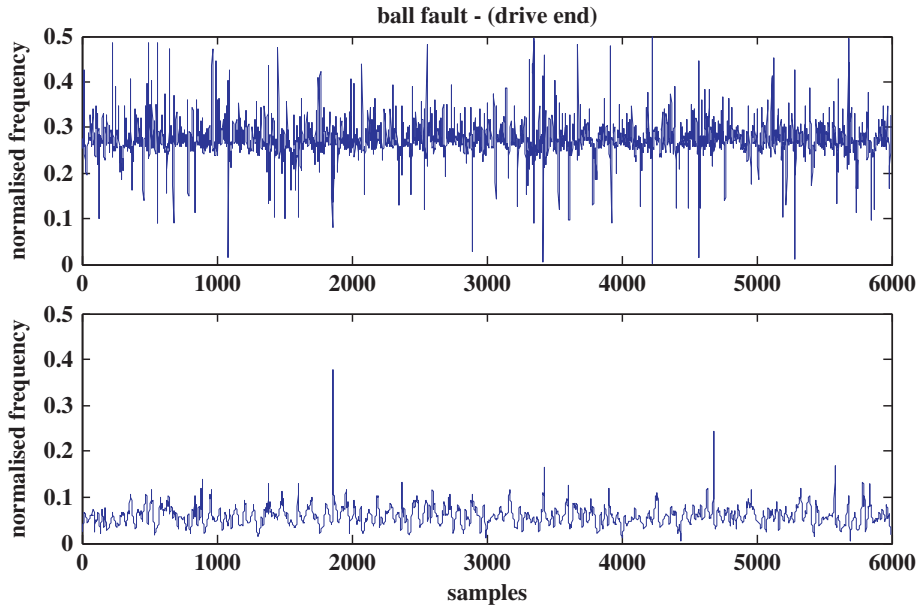

Fig. 9. Instantaneous frequency for the 2 IMFs depicted in Fig. 7, for a bearing with a 0.007 inch ball fault, 2 hp loading located at the drive end.

One of the reasons that the majority voting is among the most commonly used combination methods (apart from its simplicity) is because of the Condorcet Jury Theorem (1785) [33] (named after Marie Jean Antoine Nicolas de Caritat, marquis de Condorect 1743–1794 who wrote the "Essay on the Application of Analysis to the Probability of Majority Decisions" in 1785) that states that in the context of a jury of voters who need to make a binary decision (i.e. convict or not) and if each voter has a probability of p of being correct and P the probability of the majority of voters of being correct then:

1. $p > 0.5$ implies $P > p$ and
2. P approaches 1, for all $p > 0.5$ as the number of voters approaches infinity.

Even though the above formulation is a bit restrictive, experimental results verify its general validity making its use a very appealing option during the construction of ensemble methods.

As it was mentioned in the introduction, ensemble methods can be very effective and are among the most competitive classification methods [34] reducing the generalization error of the individual classifiers [33]. This is attributed by some researchers to the phenomenon that various types of classifiers have different "inductive biases" [50,33]. In the case of one

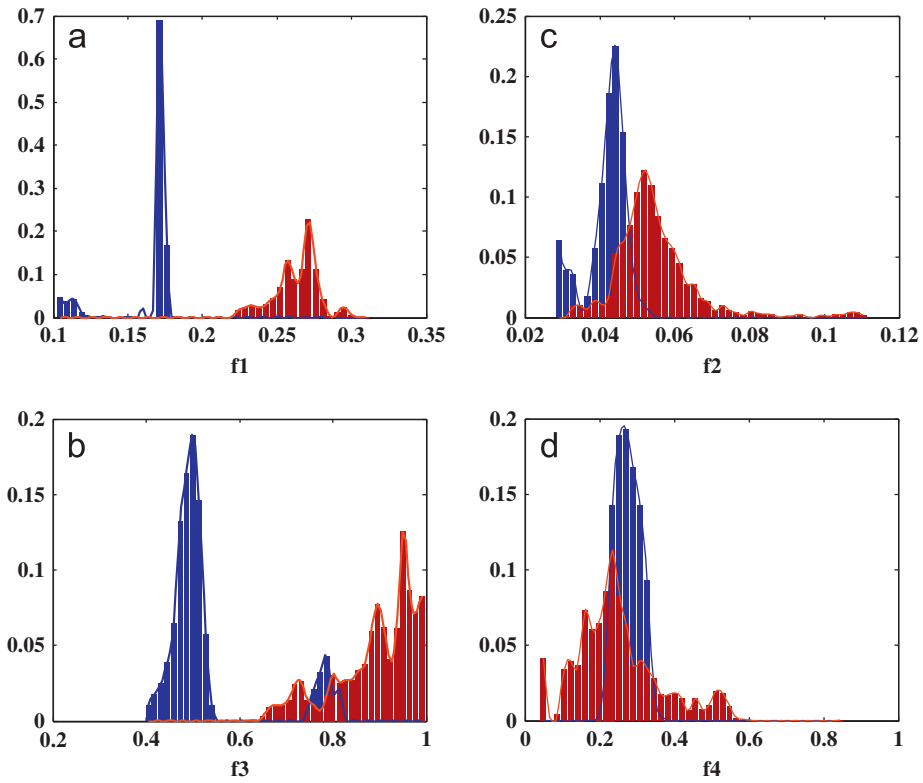

Fig. 10. Histograms for the four extracted features a) f1-median of the instantaneous frequency of the first IMF, b) f2-median of the instantaneous frequency of the third IMF, c) f3-standard deviation of the first IMF, d) f4-standard deviation of the third IMF, for normal (blue) and faulty bearings (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

class classifiers, experimental results also suggest that combining more than one classifiers can lead to an increase in accuracy [51–53].

In this work instead of using only one set of features (a set containing all four features extracted during the feature extraction stage) and combine the output of the one-class classifiers we proceeded using a form of attribute bagging [54] or feature bagging [55], also known as subset feature approach [56] or feature set partitioning [32,33]. The basic idea for using this kind of approach is to give each (one-class) classifier (detector in our case) a different "projection" of the training set. The use of different features leads to "independent" classifiers which in turn can lead to optimal improvements in accuracy, outperforming data partitioning methods (e.g. bagging and boosting) [54]. From another perspective, training on different feature sets increases diversity and diversity is among the most important factors for the success of an ensemble classifier [34]. Moreover for the case of one-class classifiers, by using fewer features and as a result, simpler models, we can avoid over-fitting the training data of the available class [52]. Apart from usually increasing the overall accuracy of the system, such decomposition facilitates faster training in the case of high dimensional feature/attribute spaces.

In this study due to the very compact feature space (only four features were extracted) we performed an almost exhaustive investigation of the possible feature subsets without employing any sort of randomization process or any other optimization algorithm [32–35,52]. More specifically we created, apart from the original feature set S1={f1, f2, f3, f4}, ten more feature (sub)sets with three and two features each:

S2={f1, f2, f3}, S3={f1, f2, f4}, S4={f1, f3, f4}, S5={f2, f3, f4} and S6={f1, f2}, S7={f3, f4}, S8={f1, f3}, S9={f2, f4}, S10={f1, f4}, S11={f2, f3}

The use of this feature/attribute bagging approach as well as the use of different detectors was performed with the hope that all the complementary information would produce a better detection performance.

5. Results

In order to test the proposed approach, we utilized a random sub-sampling approach [57]; we selected randomly 70% of the normal data for training each of the anomaly detector and we used the rest 30% of the normal data as well as the total number of faulty data for estimating the performance of the detector and repeated the whole procedure 1000 times. We tested a variety of combinations and the results are summarized in Table 2. Apart from the performance of the individual detectors acting upon each one of the 11 feature sets, the proposed ensemble detectors using different feature sets as inputs and combining the votes of the aforementioned three types of detectors, as well as ensemble detector which utilize the same type of detector at the final stage are also presented for comparison reasons. All the individual anomaly detectors were implemented using the Data Description toolbox [38].

Table 2
Detection results for individual detectors and ensemble detection schemes.

	TP rate	TN rate
S1-PCA	98.50	100
S1-Gaussian	98.59	100
S1-NN	98.63	98.66
S2-PCA	98.80	100
S2-Gaussian	98.55	100
S2-NN	98.74	96.10
S3-PCA	98.54	97.15
S3-Gaussian	97.98	100
S3-NN	98.70	100
S4-PCA	98.50	100
S4-Gaussian	98.69	100
S4-NN	98.67	96.70
S5-PCA	98.54	63.09
S5-Gaussian	98.50	97.89
S5-NN	98.59	72.68
S6-PCA	98.75	67.12
S6-Gaussian	98.56	99.62
S6-NN	98.62	100
S7-PCA	98.55	59.74
S7-Gaussian	98.48	93.56
S7-NN	98.64	63.08
S8-PCA	98.79	100
S8-Gaussian	98.70	100
S8-NN	98.59	86.47
S9-PCA	98.39	35.84
S9-Gaussian	98.47	66.21
S9-NN	98.73	43.86
S10-PCA	98.49	98.07
S10-Gaussian	98.48	100
S10-NN	98.80	99.70
S11-PCA	98.69	96.34
S11-Gaussian	98.70	89.38
S11-NN	98.73	88.41
{S2, S3, S4, S5}–PCA	98.55	100
{S2, S3, S4, S5}–Gaussian	97.78	100

Table 2 (*continued*)

	TP rate	TN rate
{S2, S3, S4, S5}–NN	98.86	99.68
{S6, S7, S8, S9, S10, S11}–PCA	∼100	100
{S6, S7, S8, S9, S10, S11}–Gaussian	98.69	100
{S6, S7, S8, S9, S10, S11}–NN	99.99	98.26
S1-{PCA, Gaussian, NN}	99.11	100
{S2, S3, S4, S5}–{PCA, Gaussian, NN}	99.94	100
{S6, S7, S8, S9, S10, S11}–{PCA, Gaussian, NN}	100	99.62
{S1, S2, S3, S4, S5}–{PCA, Gaussian, NN}	99.94	100
{S1, S6, S7, S8, S9, S10, S11}–{PCA, Gaussian, NN}	100	99.93
{S2, S3, S4, S5, S6, S7, S8, S9, S10, S11}–PCA, Gaussian, NN}	100	100
{S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11}–{PCA, Gaussian, NN}	100	100

Note: ∼100 imply that this number is the result of rounding (i.e. 83,999/84,000).

Following the convention proposed in [58] TP (true positive) rate corresponds to the detection of normal conditions and TN (true negative) rate to the detection of faulty conditions and are calculated as follows:

$$\text{TP rate} = \frac{\text{correctly detected normal objects}}{\text{total number of normal objects}}$$

$$\text{TN rate} = \frac{\text{correctly detected faulty objects}}{\text{total number of faulty objects}}$$

As it can be seen from Table 2 the hybrid approaches involving multiple feature sets along with multiple anomaly detectors perform the best, achieving in almost every case ∼100% detection accuracy both for the healthy and the faulty cases. Especially when all the 2 and 3 features sets are involved, the detection accuracy is perfect. Two of the individual detectors (PCA, Gaussian) when trained with the original four-feature set were able to detect all anomalous situations (TN rate=100%) with also very high (but not perfect) detection of the normal data (the TP rate is close to the 99% rate that is imposed during the construction of the anomaly detector (Section 3)), with the third one (NN) performing slightly worse. In the case of individual detectors and single sets with three features the Gaussian detectors have in almost all cases a very high TN rate (in three out of four sets equal to 100%) and a TP rate close to 98% exhibiting high stability. On the other hand the PCA and NN detectors are more affected by the inclusion of the less promising combination of features coming from the third IMF (see Fig. 10). In the case of individual sets with two features and individual detectors the performance in some cases deteriorates significantly.

The merit of combining more than one anomaly detectors can be justified by looking at the performance of the method when it is fed by the original four-feature set (S1 —{PCA, Gaussian, NN} in Table 2) where the detection of the faulty class remains 100%; but the TP rate exceeds the achieved performance by any of the individual anomaly detectors. This performance is raised to 100% once the attribute bagging procedure is employed with the addition of the 2 and 3 member feature sets. On the other hand, the merit of combining multiple feature sets even with a single detection algorithm can be seen more clearly in the case of the use PCA detector along with multiple two-member feature sets ({S6, S7, S8, S9, S10, S11}—PCA). In this case, even though four out of six of the individual detectors of the ensemble perform either mediocrely or bad (S9PCA), the ensemble has a remarkable 100% TN rate and an almost perfect ∼100% TP rate (with only one false alarm out of 84,000).

Among the individual detectors the Gaussian seems to be the more consistent one, even though the training data do not follow a Gaussian distribution, performing really bad (TN rate equal to 66.21) in the case of the S9 feature set, a feature set though that all detectors had problem in detecting the fault with. The NN and PCA detectors for some configurations of the input space, especially for sets with 2 members perform quite badly. Nevertheless these individual shortcomings are compensated when an ensemble scheme is employed. Finally, it is interesting to note that the verification for the combination of the two best individual features do not necessarily form the best two-set feature set [59]. In our case the two best individual features are those related to the median frequency, S8={f1, f3}, however as it can be seen in Table 2 the combination S10-NN seems to do better than S8-NN. Actually, S10 in general seems to have comparable results to

S8 when the PCA and the Gaussian anomaly detectors are involved, even though S10 contains f4 which seems to be the worst individual feature.

6. Conclusions

In this research study, we proposed an integrated and simple anomaly detection approach for the condition monitoring problem of bearings based on a compact feature set extracted from only two of the IMFs of the vibration signal along with a hybrid ensemble of one-class classifiers. It is the first time that such a hybrid ensemble method is applied in the field of bearing fault detection. The proposed method was tested on a benchmark data set including all provided load conditions (no load, 1, 2 and 3 hp), all fault locations (inner race fault, outer race fault, ball fault) and various diameters of the size fault (0.007–0.028 in.), for both the drive end and the fan end bearing. The obtained results show that the method can detect all faulty situations without false alarms by using an attribute bagging scheme and three simple detectors that require the user to provide a minimum number of free parameters (practically only the threshold of the cumulative percent variance for the PCA has to be set-with a recommended value of 90%). The minimum number of required parameters makes its application quite easy whereas the simplicity of both the involved detectors and their combination scheme make it appealing for online applications with the main computational burden residing on the extraction of the IMFs.

The set of features exploits the information carried by the two IMFs which exhibit the highest kurtosis values (in other words the two IMFs with the most "impulsive" nature) using simple measures of the instantaneous frequency "level" and the spread of each one of the two IMFs. It is interesting that although these features were extracted by inspection of normal bearings and bearings with a ball fault, they were perfectly suitable also for the detection of the inner and the outer race faults. On the other hand, we must note that these specific features might not be adequate for further separation between the different types of faults, but this is beyond the scope of this study.

If a more powerful anomaly detector was used, such as the Support Vector Data descriptor [28] the four-feature set would probably suffice at the expense of the need for parameter tuning and much greater computational cost. As it is proved here in the case of the simpler detectors each one of them lacks the capability to take full advantage of the information contained in the four-feature set. However their ensemble manages to have perfect detection performance.

In our method, we used an almost "exhaustive" approach for the creation of the feature sets without resorting to more advanced subset selection approaches [52]. The nature of the problem (small input space) along with the discrimination power of the extracted features probably does not require the use of more advanced attribute bagging methods for feature selection.

In future work, we will focus on deriving severity indices that will quantify the degradation of the bearing, something which could potentially be used for the prediction of the remaining useful life of the component. Moreover we will try to perform fault isolation/identification using a model based approach in order to avoid the dependency on faulty data.

Acknowledgements

The authors would like to thank the anonymous reviewers for their comments and suggestions Dr. George Nikolakopoulos for the fruitful discussions during the preparation of the manuscript as well as Dr. Flandrin and Dr. Rilling for their feedback regarding the use of the EMD toolbox and the selection of the stopping criterion.

References

[1] P. Zhang, Y. Du, T.G. Habetler, B. Lu, A survey of condition monitoring and protection methods for medium voltage induction motors, IEEE Trans. Energy Convers. 47 (1) (2011) 34–46.
[2] J. Antoni, R.B. Randall, Differential diagnosis of gear and bearing faults, J. Vib. Acoust. 124 (2) (2002) 165–171.
[3] T. Loutas, V. Kostopoulos, Utilising the wavelet transform in condition-based maintenance: a review with applications, advances in wavelet theory and their applications in engineering, Phys. Technol. (2012). last accessed 08.29〈http://www.intechopen.com/books/advances-in-wavelet-theory-and-their-applications-in-engineering-physics-and-technology/utilising-the-wavelet-transform-in-condition-based-maintenance〉.
[4] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine, Mech. Syst. Sig. Process. 21 (7) (2007) 2933–2945.
[5] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics, J. Sound Vib. 302 (4-5) (2007) 951–961.
[6] N.E. Huang, Z. Shen, S.R. Long, M.L.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. London, Ser. A 454 (1998) 903–995.
[7] Y. Lei, Z. He, Y. Zi, Application of an intelligent classification method to mechanical fault diagnosis, Expert Syst. Appl. 36 (6) (2009) 9941–9948.
[8] Y. Lei, Z. He, Y. Zi, Application of the EEMD method to rotor fault diagnosis of rotating machinery, Mech. Syst. Sig. Process. 23 (2009) 1327–1338.
[9] D. Yu, J. Cheng, Y. Yang, Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings, Mech. Syst. Sig. Process. 19 (2005) 259–270.
[10] Y. Peng, Empirical mode decomposition based time-frequency analysis for the effective detection of tool breakage, J. Manuf. Sci. Eng. Trans. ASME 128 (2006) 154–166.
[11] W. Guo, P. Tse, Enhancing the ability of ensemble empirical mode decomposition in machine fault diagnosis, Proceedings of the Prognostics & System Health Management Conference, Macau, 2010.
[12] M. Žvokelj, S. Zupan, I. Prebil, Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis, Mech. Syst. Sig. Process. 24 (4) (2010) 1049–1067.

[13] M. Žvokelj, S. Zupan, I. Prebil, Non-linear multivariate and multiscale monitoring and signal denoising strategy using kernel principal component analysis combined with ensemble empirical mode decomposition method, Mech. Syst. Sig. Process. 25 (7) (2011) 2631–2653.
[14] G. Georgoulas, I. Tsoumas, E. Mitronikas, C.D. Stylios, A. Safacas, Rotor fault diagnosis in asynchronous machines via analysis of the start-up transient into intrinsic mode functions, Proc. ICEM (2012).
[15] G. Wang, X.Y. Chen, F.L. Qiao, Z. Wu, N.E. Huang, On intrinsic mode function, Adv. Adaptive Data Anal. 2 (03) (2010) 277–293.
[16] N.E. Huang, Z. Shen, S.R. Long, A new view of nonlinear water waves: the Hilbert Spectrum, Annu. Rev. Fluid Mech. 31 (1999) 417–457.
[17] N.E. Huang, M.L.C. Wu, S.R. Long, S.S. Shen, W. Qu, P. Gloersen, K.L. Fan, A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, Proc. R. Soc. London, Ser. A: Math., Phys. Eng. Sci. 459 (2037) (2003) 2317–2345.
[18] Z. Wu, N.E. Huang, A study of the characteristics of white noise using the empirical mode decomposition method, Proc. R. Soc. London, Ser. A: Math., Phys. Eng. Sci. 460 (2046) (2004) 1597–1611.
[19] P. Flandrin, G. Rilling, P. Goncalves, Empirical mode decomposition as a filter bank, IEEE Signal Process Lett. 11 (2) (2004) 112–114.
[20] L. Li, H.B. Ji, Signal feature extraction based on an improved EMD method, Measurement 42 (2009) 796–803.
[21] B. Xuan, Q. Xie, S. Peng, EMD sifting based on bandwidth, IEEE Signal Process Lett. 14 (8) (2007) 537–540.
[22] C. Junsheng, Y. Dejie, Y. Yu, Research on intrinsic mode function (IMF) criterion in EMD method, Mech. Syst. Sig. Process. 20 (2006) 817–824.
[23] G. Rilling, P. Flandrin, P. Goncalves, On empirical mode decomposition and its algorithms, Proceedings IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, NSIP03, Grado, Italy, (2003).
[24] The EMD toolbox 〈http://perso.ens-lyon.fr/patrick.flandrin/emd.html〉 (last accessed 08.29.2012).
[25] R.L. Allen, D. Mills, Signal Analysis: Time, Frequency, Scale, and Structure, John Wiley and Sons Inc, Hoboken, NJ 07030, USA, 2004.
[26] M. Markou, S. Singh, Novelty detection: a review-part 1: statistical approaches, Sig. Process. 83 (12) (2003) 2481–2497.
[27] M. Markou, S. Singh, Novelty detection: a review-part 2: neural network based approaches, Sig. Process. 83 (12) (2003) 2499–2521.
[28] D.M.J. Tax, One-Class Classiffication; Concept-Learning in the Absence of Counter-Examples. Ph.D. Thesis. Delft University of Technology (2001).
[29] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (3) (2009) 1–58.
[30] T.S. Khawaja, A Bayesian Least Squares Support Vector Machines Framework for Fault Diagnosis and Failure Prognosis, Ph.D. Thesis. Georgia Institute of Technology (2010).
[31] R.O. Duda, P.H. Hart, D.G. Stork, Pattern Classiffcation, Wiley-Interscience, New York, 2000.
[32] L. Rokach, Ensembled-based classifiers, Arttificial Intell. Rev. 33 (2010) 1–39.
[33] L. Rokach, Pattern Classification Using Ensemble Methods, World Scientific Publishing Company Incorporated, 2010.
[34] L.I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Willey-IEEE, 2004.
[35] L. Rokach, Taxonomy for characterizing ensemble methods in classification tasks: a review and annotated bibliography, Comput. Stat. Data Anal. 53 (12) (2009) 4046–4072.
[36] R. Polikar, Ensemble based systems in decision making, IEEE Circuits Syst. Mag. 6 (3) (2006) 21–45.
[37] L. Tarassenko, P. Hayton, M. Brady, Novelty detection for the identification of masses in mammograms, Proceedings of the Fourth International Conference on Artificial Neural Networks 409 (1995) 442–447.
[38] The Data Description Toolbox, 〈http://homepage.tudelft.nl/n9d04/dd_tools.html〉 (last accessed 08.29.2012).
[39] V. Vapnik, Statistical Learning Theory, Wiley, 1998.
[40] V. Cherkassky, F. Mulier, Learning From Data: Concepts, Theory, and Methods, John Wiley & Sons, 2007.
[41] S. Haykin, Neural Networks and Learning Machines, third ed. Prentice Hall, 2008.
[42] C. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.
[43] I. Jolliffe, Principal Component Analysis, John Wiley & Sons, 2005.
[44] P. Taylor, Statistical Methods, in: M.R. Berthold, D.J. Hand (Eds.), Intelligent Data Analysis: An Introduction, Springer, 2007.
[45] K.A. Loparo, Bearing Vibration Dataset, Case Western Reserve University 〈http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website〉 (last accessed 08.29.2012).
[46] X. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy Inference, Mech. Syst. Sig. Process. 18 (5) (2004) 1077–1095.
[47] J.P.Marques de Sa, Pattern Recognition: Concepts, Methods and Applications, Springer, 2001.
[48] J. Astola, P. Kuosmanen, Fundamentals of Nonlinear Digital Filtering, CRC, Boca Raton, FL, 1997.
[49] D. Bahler, L. Navarro, Methods for combining heterogeneous sets of classifiers, Proceedings 17th National Conference on Artificial Intelligence (AAAI), Workshop on New Research Problems for Machine Learning (2000).
[50] S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma, Neural Comput. 4 (1995) 1–58.
[51] L. Nanni, Experimental comparison of one-class classifiers for online signature verification, Neurocomputing 69 (7) (2006) 869–873.
[52] V. Cheplygina, D.M.J. Tax, Pruned random subspace method for one-class classifiers, Lect. Notes Comput. Sci. 6713 (2011) 96–105.
[53] D.M.J. Tax, R. Duin, Combining one-class classifiers, Multiple Classifier Syst. 2001 (2001) 299–308.
[54] R. Bryll, R. Gutierrez-Osuna, F. Quek, Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets, Pattern Recognit. 36 (2003) 1291–1302.
[55] A. Lazarevic, V. Kumar, Feature bagging for outlier detection, Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, (2005), pp. 157–166.
[56] M.A. Aly, A.F. Atiya, Novel methods for the feature subset ensembles approach, Int. J. Artificial Intell. Mach. Learn. 6 (4) (2006).
[57] N. Japkowicz, M. Shah, Evaluating Learning Algorithms; A Classification Perspective, Cambridge University Press, 2011.
[58] J. Zhang, Q. Yan, Y. Zhang, Z. Huang, Novel fault class detection based on novelty detection methods, Proc., Intell. Comput. Sig. Proc. Pattern Recog. 345 (2006) 982–987.
[59] T.M. Cover, The best two independent measurements are not the two best, IEEE Trans. Syst. Man Cybern. 1 (1974) 116–117.